

Proceedings of the ECML/PKDD – 2003

# First European Web Mining Forum

Edited by

Bettina Berendt, Andreas Hotho, Dunja Mladenic,  
Maarten van Someren, Myra Spiliopoulou, and Gerd Stumme



14<sup>th</sup> European Conference on Machine Learning  
7<sup>th</sup> European Conference on Principles and Practice of Knowledge  
Discovery in Databases

22 – 26 September 2003  
Cavtat - Dubrovnik  
Croatia

ISBN: 953-6690-31-4



© IRB 2003

Proceedings of the ECML/PKDD – 2003  
**First European Web Mining Forum**

Edited by  
Bettina Berendt, Andreas Hotho, Dunja Mladenic,  
Maarten van Someren, Myra Spiliopoulou, and Gerd Stumme

Workshop Chairs: Stefan Kramer, Technische Universität München  
Luis Torgo, University of Porto

Publisher: Ruđer Bošković Institute, Bijenička cesta 54, P.O.B. 180, 10002 Zagreb, Croatia

**ISBN:** 953-6690-31-4

Zagreb 2003

## **ECML/PKDD-2003: 1st European Web Mining Forum [EWMF 2003]**

The Web presents a key driving force for a large spectrum of applications in which a user interacts with a company, a governmental authority, a non-governmental organization, or another non-profit institution. The application providers should combine knowledge about user expectations and Web semantics, in order to form personalised, user-friendly, and business-optimal services. Enabling methodologies include data mining, text mining, and ontology learning. Recipient technologies include user profiling, usage analysis, ontology extraction for the Semantic Web, intelligent search and recommendation systems based on user preferences, page content, and site semantics.

The EWMF'03 workshop is organised by the KDNet interest group Web Mining Forum on knowledge discovery from and about the Web. It aims to bring together various perspectives on Web mining and stress the synergy effects between Web usage mining, Web content mining and Web intelligence, and of Semantic Web Mining.

We would like to take this opportunity to thank the members of the Program Committee and the secondary reviewers for their great assistance in reviewing the submitted publications.

By Bettina Berendt, Andreas Hotho, Dunja Mladenic, Maarten van Someren, Myra Spiliopoulou, Gerd Stumme, July 2003

### **Organizing Committee**

- Bettina Berendt  
Institute of Information Systems, Humboldt University Berlin,  
email: berendt@wiwi.hu-berlin.de
- Andreas Hotho  
Institute AIFB, Universität Karlsruhe, Germany  
email: hotho@aifb.uni-karlsruhe.de
- Dunja Mladenic  
J. Stefan Institute, Department for Intelligent Systems, Ljubljana, Slovenia  
email: Dunja.Mladenic@ijs.si
- Maarten van Someren  
SWI, University of Amsterdam, The Netherlands  
email: maarten@swi.psy.uva.nl
- Myra Spiliopoulou  
Institute of Technical and Business Information Systems, Otto-von-Guericke-  
Universitaet Magdeburg  
email: myra@iti.cs.uni-magdeburg.de
- Gerd Stumme  
Institute AIFB, Universität Karlsruhe, Germany  
email: stumme@aifb.uni-karlsruhe.de

## Program Committee

- Ed Chi (Xerox Parc, Palo Alto, USA)
- Ronen Feldman (Bar-Ilan University, Ramat Gan, Israel)
- Marko Grobelnik (J. Stefan Institute, Ljubljana, Slovenia)
- Oliver Günther (Humboldt University Berlin, Germany)
- Stefan Haustein (Universität Dortmund, Germany)
- Jörg-Uwe Kietz (kdlabs AG, Zuerich, Switzerland)
- Ee-Peng Lim (Nanyang Technological University, Singapore)
- Alexander Maedche (Robert Bosch GmbH, Stuttgart, Germany)
- Brij Masand (Data Miners, Boston, USA)
- Yannis Manolopoulos (Aristotle University, Greece)
- Ryszard S. Michalski (George Mason University, USA)
- Bamshad Mobasher (DePaul University, Chicago, USA)
- Claire Nedellec (Université Paris Sud, France)
- George Paliouras (Demokritos National Centre for Scientific Research, Athens, Greece)
- Jian Pei (Simon Fraser University, Canada)
- John R. Punin (Rensselaer Polytechnic Institute, Troy, N.Y., USA)
- Jaideep Srivastava (University of Minnesota, Minneapolis, USA)
- Rudi Studer (Universität Karlsruhe, Germany)
- Stefan Wrobel (Fraunhofer Institute for Autonomous Intelligent Systems, Sankt Augustin, Germany)
- Mohammed Zaki (Rensselaer Polytechnic Institute, USA)
- Osmar Zaiane (University of Alberta, Edmonton, Alberta / Canada)

## Secondary reviewers

- Prasanna Desikan (University of Minnesota, Minneapolis, USA)
- Haiyang Liu (University of Minnesota, Minneapolis, USA)
- Oliver Guenther (Humboldt University Berlin, Germany)

This Workshop is partially supported by the European FP5 FET project "European Knowledge Discovery Network of Excellence (KDNNet)".

# Table of contents

## Organising Web Content

Mining Web sites using wrapper induction, named entities and post-processing, Georgios Sigletos, Georgios Paliouras, Constantine D. Spyropoulos, Michalis Hatzopoulos .....	1
Semantic Outlier Analysis for Sessionizing Web Logs, Jason J. Jung, Geun-Sik Jo .....	13
Construction of Web Community Directories using Document Clustering and Web Usage Mining, Dimitrios Pierrakos, Georgios Paliouras, Christos Papatheodorou, Vangelis Karkaletsis, Yannis Ioannidis, Marios Dikaiakos .....	25
Language Models for Intelligent Search Using Multinomial PCA, Wray Buntine, Petri Myllymaki, Sami Perttu .....	37

## Analysing user behaviour

Evaluation and Validation of Two Approaches to User Profiling, F. Esposito, G. Semeraro, S. Ferilli, M. Degemmis, N. Di Mauro, T.M.A. Basile, and P. Lops .....	51
Monitoring the Evolution of Web Usage Patterns, Steffan Baron, Myra Spiliopoulou .....	64

## Architectures for Web Mining

Beyond user clicks: an algorithm and an agent-based architecture to find out user behaviour, E. Menasalvas, S. Millán, M. Pérez, E. Hochsztain, V. Robles, O. Marbán, J. Peña, A. Tasistro .....	80
Greedy recommending is not always optimal, M. van Someren, V. Hollink and S. ten Hagen .....	93